7755 2819 E

DIAGNOSTIC RULES GENERATOR

Phase I

Final Report

Prepared for:

USAF School of Aerospace Medicine
Human Systems Division (AFSC)
United States Air Force
Brooks AFB, Texas 78235-5301

Prepared by:

Dr. P.R. Saunders
Strategic Analysis Division
Analytics
3702 Pender Drive, Suite 431
Fairfax, Virginia 22030

31 UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

1b. RESTRICTIVE MARKINGS
N/A

2a. SECURITY CLASSIFICATION AUTHORITY
N/A

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE
N/A

3. DISTRIBUTION / AVAILABILITY OF REPORT
Unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S)
N/A

5. MONITORING ORGANIZATION REPORT NUMBER(S)
FY7624

6a. NAME OF PERFORMING ORGANIZATION
ANALYTICS

6b. OFFICE SYMBOL *(If applicable)*
N/A

7a. NAME OF MONITORING ORGANIZATION
USAFSAM/NG

6c. ADDRESS *(City, State, and ZIP Code)*
3702 Pender Drive, Suite 431
Fairfax, Virginia 22030

7b. ADDRESS *(City, State, and ZIP Code)*
Brooks AFB, Texas 78235-5301

8a. NAME OF FUNDING / SPONSORING ORGANIZATION
USAFSAM/NG

8b. OFFICE SYMBOL *(If applicable)*

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
F33615-87-C-0621

8c. ADDRESS *(City, State, and ZIP Code)*
Brooks AFB, Texas 78235-5301

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO	WORK UNIT ACCESSION NO.
87-088			

11. TITLE *(Include Security Classification)*

12. PERSONAL AUTHOR(S)
Dr. P. R. Saunders

13a. TYPE OF REPORT
Final

13b. TIME COVERED
FROM 3Sep87 TO 3Mar88

14. DATE OF REPORT *(Year, Month, Day)*
3 March 1988

15. PAGE COUNT

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
Artificial Intelligence in Medicine
Machine Learning
Thallium Myocardial Scintigraphy

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

This report investigates the feasibility of the automated learning of physician's diagnostic criteria for medical telemetry data from a set of examples. Using thallium myocardial scintigraphy as an illustrative domain, an architecture is established for a Diagnostic Rules Generator, and the required machine learning capability is prototyped and evaluated.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT
☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL
Dr. P. R. Saunders

22b. TELEPHONE *(Include Area Code)*
(703) 246-9060

22c. OFFICE SYMBOL

DD Form 1473, JUN 86 *Previous editions are obsolete.*

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

FIGURES & TABLES

ACKNOWLEDGMENTS

I wish to thank the following individuals for their support:

Dr. William Clardy
Dr. Bryce Hartman
Dr. Londe Richardson
Mr. Tom Kay

Rin Saunders
Principal Investigator

DIAGNOSTIC
RULES GENERATOR
PHASE I FINAL REPORT

1.0 INTRODUCTION

This report investigates the feasibility of a new medical application of artificial intelligence (AI): a computer system which learns a physician's diagnostic criteria for medical telemetry data from a set of examples. The system will be called the Diagnostic Rules Generator (DRG).

Diagnostic expert systems and their close relatives, advisory systems, have amply demonstrated their usefulness in medicine. The diagnostic skill level of the classic MYCIN system compares favorably to that of Stanford infectious disease specialists [YU84]. MYCIN's descendants are available to physicians as commercial products. Advisory systems like ONCOCIN [SHO84] (oncology protocols) and ATTENDING [MIL84] (anesthesiological procedures) are in routine use.

Medical telemetry is potentially a high-payoff domain for expert systems. For chest films, Garland and others have shown that radiologists routinely miss about 30% of abnormalities. Telemetry interpretation is presumably even more difficult because it is more subjective. Techniques like electrocardiography, vectorcardiography, echocardiography, electrogastrograms and scintiligraphy result in displays that are more abstract and less representational than, say, an X-ray. Expertise plays a larger role in interpretation. And expertise is scarcer--more physicians can read an X-ray than can read a scintiligram.

Building a medical expert system requires a medical knowledge base and an expert system shell. Creating the knowledge base is serious bottleneck. Knowledge bases are notoriously difficult and time-consuming to build and validate. And the time required is the expert's--whose limited availability motivated the building of an expert system.

The goal of DRG is to overcome the bottleneck by producing knowledge bases more or less automatically, learning the rules from examples. The system would require an expert's assistance only when it found an ambiguity in the data requiring clarification.

The impetus for DRG arose from a 20-year longitudinal study of cardiovascular disease conducted by the United States Air Force School of Aerospace Medicine (USAFSAM). The study uses planar thallium-201 scintigraphy as a means of detecting asymptomatic coronary artery disease. For conciseness, the scintigraphy technique and its products will be referred to as "thallium" and "thallium imagery".

DRG was designed using thallium imagery as model for telemetry data. After it has been successfully applied in that domain, it will be extended to cover other kinds of medical telemetry.

Within that context, DRG would offer the following benefits:

- o LEVERAGING EXPERTISE. Expert systems enable users to perform at the highest available level of expertise--that of the expert(s) whose knowledge underlies the system.

- CONSERVING EXPERTISE. . A formally trained interpreter of thallium imagery will normally have a 2-3 year nuclear medicine fellowship involving 6 months or more in cardiology, or a cardiology fellowship with 200 hours of physics and 500 hours of clinical science. At times, USAFSAM must use physicians without that specialized training to interpret imagery. These individuals become experts on the job--only to leave. (In fact, the amount of time a doctor has spent on the job can be calculated from the percentage of thallium-based diagnoses confirmed by arteriography [KAY88].) A DRG-built knowledge base would preserve their learning and transmit it to the physician's replacement.

- OBJECTIVIZING EXPERTISE. A DRG-built rule base would enable experts to combine their knowledge by providing an objective set of criteria to discuss and evaluate. The result could be a better knowledge base than could be obtained from a single diagnostician.

- CONSISTENT AND REPRODUCIBLE DIAGNOSES. Although thallium interpretation is somewhat subjective, precise numbers underlie the image. A rule base would apply a definite diagnostic standard that would be consistent and reproducible across both doctors and patients.

Even without an expert system, DRG would offer the following benefit:

- RECONSTRUCTING EXPERTISE. Although most patients remain in the study for 20 years, doctors often leave after 3-4 years. Since thallium interpretation has subjective component, a doctor reviewing the work of a long-departed colleague may be at a loss to understand why certain diagnostic decisions were made. DRG would address the problem by presenting diagnostic rules in English which explain the departed doctor's decisions. DRG would also write the rules in a form readable by an expert system shell. One could build a library of "doctors-on-a-disk" and compare how various colleagues would have interpreted an image.

The feasibility study for DRG followed the following plan:

1). An architecture was developed for DRG.

2). Areas of significant technology risk were identified in the architecture. The only significant risk area proved to be machine learning from examples, which is still a laboratory research field in AI.

3). The demands were identified that USAFSAM's application-- learning diagnostic rules for medical telemetry data-- would place on the machine learning module.

4). The literature in machine learning was reviewed in search of a system that would meet these demands. None was found, so:

5). The feasibility of DRG was demonstrated by successfully designing and prototyping a machine learning system that will meet the requirements.

The study was conducted under Phase I of a Small Business Innovative Research (SBIR) program sponsored by the Human Systems Division of USAFSAM. The full DRG system will, at the sponsor's option, be built under a Phase II effort.

1.1 Organization of the Report

Section 2 provides a brief overview of how thallium imagery is created and interpreted. The following section describes an architecture for machine learning of diagnostic criteria from telemetry and identifies areas of technology risk. The only significant technology risk is the machine learning algorithm. Section 4 describes the prototype learning algorithm in detail. The final section evaluates the algorithm by applying it to sample problems.

2.0 AN OVERVIEW OF PLANAR THALLIUM SCINTIGRAPHY AND ITS INTERPRETATION

USAFSAM uses thallium to screen for coronary artery disease (CAD) in asymptomatic patients [SCH87]. Suspicion of CAD may be raised by abnormal results on a stress EKG, Holter monitor or other test, or by risk factors such cholesterol levels. Angiography is considered the definitive technique for diagnosing CAD; however, since angiography is invasive, requiring cardiac catheterization and the injection of a radiopaque dye often associated with allergic reactions, USAFSAM performs thallium imagery first. The outcome of the thallium test is graded as "normal", "borderline" or "abnormal". Patients with normal imagery are considered free of significant CAD and are not subject to angiography.

Thallium imagery reveals the presence of coronary artery disease indirectly, by depicting the perfusion of blood into the left ventricle of the heart. Three kinds of abnormalities can be visualized: reversible ischemia, irreversible ischemia, and certain anatomical defects. In the asymptomatic patients usually seen at USAFSAM, reversible ischemia is the expected positive finding.

The imaging technique is based on the fact that thallium is metabolized much like potassium. Prior to imaging, the patient is exercised on a treadmill to a predetermined peak level. As with all muscle, exercise depletes the heart of potassium. The left ventricle undergoes most of the depletion because it performs the main pumping action.

Within one minute of attaining the peak exercise level, the patient is injected with 2.2 millicuries of thallium-201 chloride through an IV line into the arm.

The left ventricle scavenges the thallium from circulation to replace potassium. If perfusion is normal, the thallium is absorbed rapidly then gradually washes out, attaining the half-way point typically in 84 minutes [GER87]. The degree of thallium absorption into regions of the left ventricle muscle wall reflect varying degrees of perfusion and reperfusion (perfusion during washout).

The patient is placed under a gamma camera and imaged within six minutes of injection. Three views are taken: the anterior (ANT), 45 degree left anterior oblique (45-LAO) and 67 degree left anterior oblique (67-LAO). The views are repeated after four hours of rest. During that time, substantial washout will have occurred from non-ischemic tissue, and reversible ischemia will have largely disappeared.

Ischemia appears as a region on an image showing less absorption than its environs (a "perfusion defect", or "cold spot"). In reversible ischemia, the ischemic muscle's thallium uptake increases as ischemia disappears. The corresponding images show a region that grows hotter over time (a "reperfusion defect"). If ischemia is irreversible, as with fibrosis due to a prior infarct, the cold spot stays cold (a "fixed perfusion defect").

Interviews and observation of experts at work show that expert interpretation of thallium images involves reasoning that uses rules, heuristics and a simple model of coronary perfusion. Rules and heuristics are reasonably well-understood methods for representing knowledge. Causal modeling is a new but rapidly developing field. All three forms of knowledge representation appear in present-day medical expert systems. Notable for its absence is analogic reasoning, which was initially expected to play some part in interpretation. That is a comforting finding from a feasibility standpoint, since automated analogic reasoning is in its infancy.

Representative rules and heuristics underlying expert feature extraction and interpretation are given below.

- A case is graded as abnormal if there are one or more perfusion defects in the exercise images with matching reperfusion defects in the rest images.
- A case is graded as borderline if there is an unmatched defect of modest size.
- A small reperfusion defect can be a normal finding.
- Prior infarcts and anatomical defects are separate findings which do not affect the grading into normal/borderline/abnormal categories.
- A change in pixel intensity within a region must be at least two standard deviations to be significant. In the MICAS system, color bands are used to group pixels into one-standard-deviation bands; hence regions with significant differences will appear in different colors. In close cases, the numerical pixel values are examined.
- The papillary muscles (which lie inside the ventricle and close the mitral valve) and in female patients, the breasts may attenuate the gamma rays and cause cold spots.
- The lungs, liver and spleen may absorb thallium and contribute to the background. The technician can adjust the patient's position to move other organs out of the field of view, except for the lungs, which will necessarily fall within the image.
- Washout may cause the muscle walls to appear thicker on the first set of images.
- A change in muscle wall thickness that involves more than half the thickness of the wall is always an abnormal finding. Smaller changes may be artifacts due to absorption of thallium by intervening muscle tissue.
- Apparent defects located in the 1/5 of the image nearest the valve plane should be discounted. Thallium uptake is variable near the valve plane.
- Apical thinning is a normal variant which is difficult to discriminate from a cold spot. This is especially true of small hearts.
- The papillary muscles can absorb thallium and create a hot spots, especially in the ANT and anteriolateral views. This is important because a hot spot can bias image pre- processing.
- Apparent reverse reperfusion (a cold spot that appears worse on the rest image) may signify pathology [FRO87], but can be an artifact due to saturation of the gamma camera with counts.
- A rare but almost pathognomic pattern is the "reversing horseshoe", in which the exercise image shows an unusually hot valve plane while the rest image shows a cold valve plane and hot apex [KAY88].
- Three-vessel disease may cause slow global washout--the entire ventricle experiences a reperfusion defect.
- The validity of the test can be affected by patient compliance problems. Failure to fast prior to the test causes the stomach to absorb much of the thallium-- physicians

detect this problem both on imagery and at the O-Club. Exercising prior the rest imagery can cause the heart to completely reperfuse.

- o Test validity can also be affected by active gout, diabetes, hypertension, beta blockers, vasodilators, spasm in the arm into which the thallium is injected, and defective scintillation tubes in the camera.

A simple model of circulation assists the expert in interpreting defects. The views taken at USAFSAM show eight regions of the ventricle muscle wall: anterior, posterior, apical, inferior, septal, posterolateral, inferioapical and anterolateral. These regions are also called walls. Each wall is supplied by one or more of three coronary arteries: the left anterior descending (LAD), left circumflex, and right coronary arteries. The distribution of these arteries is somewhat variable. In general, the LAD perfuses the anterior and septal walls together with the apex. Left circumflex lesions most commonly affect the apex, lateral wall and posterior wall. The right coronary artery may provide collateral circulation to the inferior wall. Reperfusion defects are caused by stenotic lesions in one or more of these arteries.

The model enables the diagnostician to assess the strength with which a set of defects implies the presence of CAD. Multiple defects within the distribution of a single artery are strongly suggestive, although multi-vessel disease is by no means infrequent. The model also predicts that in single-vessel disease, collateral circulation may obscure a defect, e.g. in the apex.

3.0 RISK ASSESSMENT OF A DESIGN FOR A DIAGNOSTIC RULES GENERATOR

The DRG architecture diagram shows the creation of a knowledge base and its exploitation by an expert system shell (figure 1).

DRG learns rules from a set of images paired with expert diagnoses, called a training set. The DRG system extracts rules from the data in four successive steps, each performed by a subsystem of DRG:

- the pre-processor makes the image more clear and distinct via a series of mathematical image processing operations;

- the feature extractor identifies image features of diagnostic significance and outputs feature descriptions;

- the learning subsystem generalizes diagnostic criteria that explain the diagnoses in the training set in terms of the diagnostic features. The expert may be consulted during this process to clarify ambiguous data. The criteria are stated in a structured, English-like format readable by humans but also having a definite meaning to DRG;

- the rule writer translates the diagnostic criteria to the format required by a specific expert system shell. It may be desirable to have one or more rule writers which can be "dropped in" to DRG to accommodate various shells.

The feasibility of DRG can be demonstrated by demonstrating the feasibility of each of the four parts. The technology risks inherent in building each subsystem are examined below.

3.1 Risk Assessment for the Pre-Processor

Without special processing, thallium imagery looks indistinct. It can require effort to distinguish basic anatomical features, let alone abnormalities. Since feature extraction requires the computerized detection of abnormal image regions, image pre-processing will considerably aid in reliable feature detection.

After a review of the computer vision literature, the principal investigator designed pre-processing methods to clarify and enhance thallium imagery. It was later determined that a virtually identical pre-processor design is implemented by MICAS imaging system at USAFSAM. Evidently the pre-processing designed to benefit the feature extraction subsystem parallels pre-processing designed to assist the human eye in performing the same task. The technology risk of this sub-system is nil: the sub-system has already been built.

Pre-processing applies an ordered sequence of mathematical operations to image data. As produced by the gamma camera, raw image data consists of a 64x64 grid of 16-bit integers. These numbers correspond to the gamma ray counts from each scintillation tube in the camera, and give the gray-scale intensity for each pixel in the image.

The planned pre-processor uses four successive mathematical operations to clarify image features. Each addresses a specific source of obfuscation in the image. The planned pre-processor is described and compared with MICAS image processing.

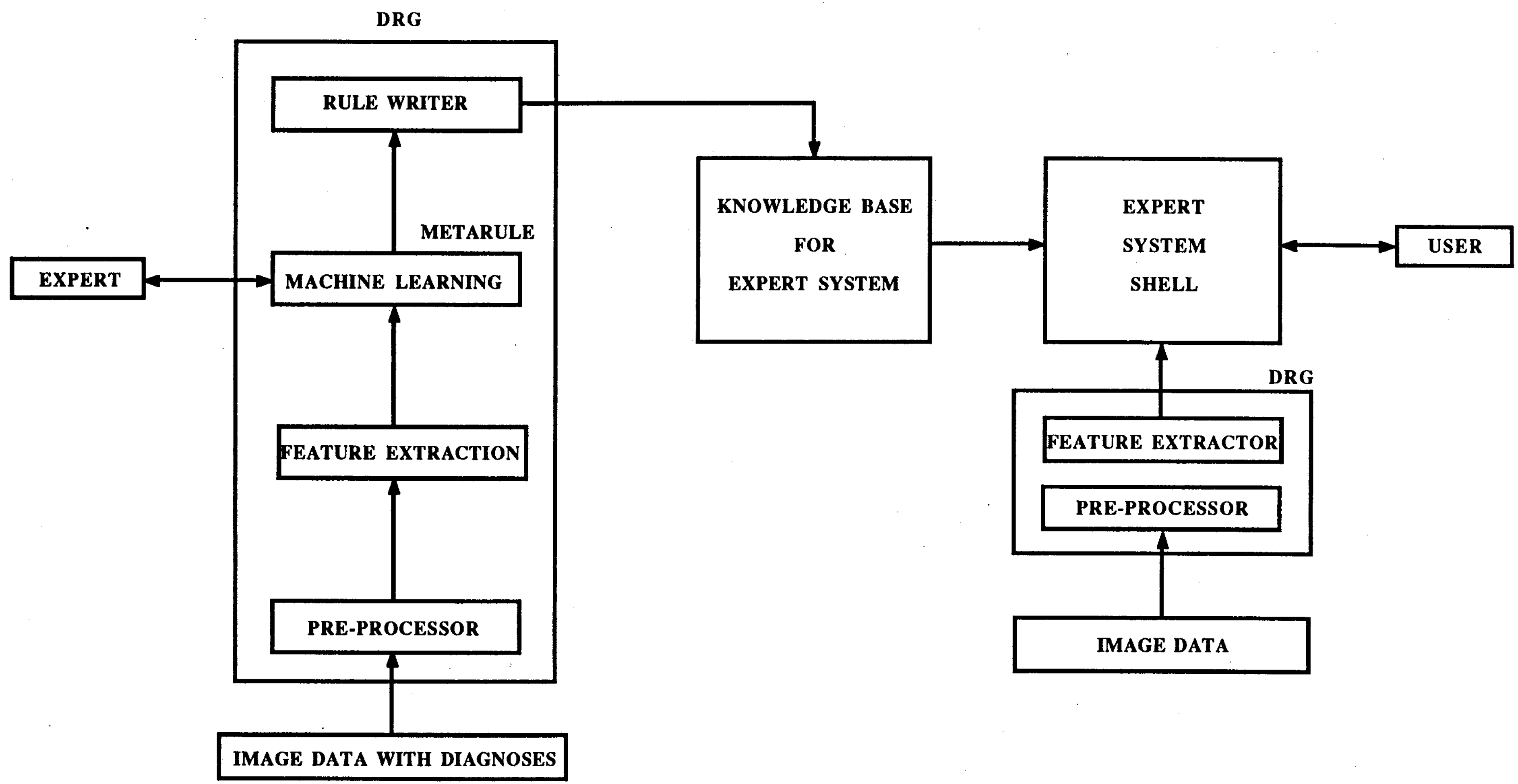

Figure 1. Architecture for the Diagnostic Rules Generator

1). SMOOTHING. Smoothing is based on the assumption that image features are large in size compared to pixels; therefore, an individual pixel value should not differ greatly from neighboring values. If a pixel seem "out of place" in its neighborhood, smoothing adjusts the value to lie closer to the average of neighboring values. MICAS uses a 5x5 convolution (a mathematical operation) to implement smoothing.

2). BACKGROUND SUBTRACTION. A raw thallium image typically contains much background noise. This operation is designed to reduce background in both the heart and non-heart portions of the image. The DRG pre-processor design specifies the bilinear subtraction method used by Watson [WAT81] for thallium image processing. MICAS's experience with thallium data has shown that subtracting from each pixel an amount equal to 22.5% of the highest pixel value yields a similar result in less computer time.

3). DISTRIBUTIONAL TRANSFORM. The human eye (and computer programs) depend on contrast to distinguish features. Image contrast can be enhanced by transforming the pixel values so that they conform to a selected statistical distribution--typically the Poisson (a close relative to the Gaussian, or normal, distribution). MICAS performs a similar operation in two steps. First, pixel values are normalized to lie between 0 and 255. Then a special distribution is imposed on the image, if it is to be displayed in black and white (the transformation is not applied for color display). The distribution is not a standard statistical one, but one specially crafted for this application.

4). EDGE ENHANCEMENT. Edge enhancement strengthens the contrast between an image feature and surrounding pixel values. The features appear more prominent in an edge-enhanced image. The first step in this process is edge detection--identifying the boundaries between regions. This is normally accomplished via a Fourier transform, a mathematical technique that allows the identification of image areas where pixel values are rapidly changing. These areas are likely to be edges. MICAS uses a fast version of the Fourier transform for edge detection, then enhances edge contrast by brightening the pixels comprising an edge.

MICAS performs two further operations not in the DRG design. After edge enhancement, it performs Watson's bilinear subtraction [WAT81]. This background subtraction technique was found to be more effective in MICAS when applied after edge enhancement. Finally pixel values that fall in the border zone--the zone outside the left ventricle as determined by edge detection--are set to zero. This eliminates both counts due to background radiation, and due to absorption of thallium by other anatomical structure, such as the lungs and liver, when those structures do not overlie the heart.

MICAS presents the processed data in several visual formats, e.g. gray-scale and color. A defect seen in any format is deemed significant. Since DRG will operate upon the underlying numbers, multiple display formats do not appear in the DRG pre-processor design.

MICAS's pre-processor will be fully adequate for the DRG pre-processing subsystem.

3.2 Risk Assessment for the Feature Extractor

Feature extraction means detecting features of diagnostic significance in an image and describing their relevant attributes.

To a layman, a pre-processed thallium image appears as a heart-shaped object with a hollow center (in the ANT view), or a blobby ring with a chunk missing (in the LAO views). An expert sees features (such as muscle walls, perfusion defects and reperfusion defects) which give meaning to the image. Diagnoses are based on features rather than pixel values per se.

Feature extraction replicates the expert's view of an image. Feature extraction is pattern matching--an expert knows in general terms what a feature looks like, and seeks specific instances in the image.

There are a number of techniques by which feature extraction can be performed.

Template matching is a simple technique. A template is a test that can be applied to an image region to tell if it is a feature. For example, "a region that grows hotter by more than two standard deviations from the mean pixel value" is a template for a reperfusion defect. Features are identified by searching the data with templates for regions that fit.

Syntactic analysis is a more sophisticated alternative. It is useful when a feature's identification depends in part on its relationship to other features. In EKG interpretation, for example, a P wave precedes a QRS complex. A wave might match the shape of a P wave template, i.e. look like a typical P wave in isolation, but be rejected as a P wave because something other than a QRS complex comes next. Syntactic analysis uses parsing techniques borrowed from natural language analysis--the possible ordering of EKG waves forms a kind of grammar, in which a P wave precedes a QRS complex just as an adjective precedes a noun.

To identify appropriate feature extraction techniques, let us reconstruct the way that an expert approaches an image.

The first extraction operation an expert performs is to detect missing walls. These are evident at a glance. Most thallium images have a shape which is characteristic of the view: heart-like or U-shaped (a ring with a missing chunk). A severe perfusion defect may cause a section of the image to disappear, giving a normally U-shaped view a J-shape, for example.

Correspondingly, the DRG feature extractor should first match the outline of the ventricle--which will be the largest continuous edge reported by the pre-processor's edge detector--against a set of geometrical templates representing normal images and images with one or more missing walls.

Artificial neural systems (ANS, or neural nets) are a pattern recognition technology that might be applied to missing wall detection and other feature extraction problems.

Neural nets can be contrasted with a template matching approach that uses curve fitting.

Curve fitting begins with a library of mathematical descriptions (templates) for shapes, such as the U-shape and the J-shape. The descriptions have adjustable parameters for location, orientation and size. An optimization technique such as gradient search is used to find the best parameters for each candidate shape; the description with the best overall fit is deemed the matching template.

Neural networks are a more general pattern recognition tool that can be "trained" to recognize a shape from a large number of examples [RUM87]. Neural networks use a technique similar to gradient search. Unlike the template approach, a neural network can learn to recognize a shape without a prior mathematical description of the shape.

The expert mentally divides the image into regions representing anatomical entities (figure 2). Once the outline of the ventricle is found by the pre-processor's edge detection technique, DRG can identify anatomical regions by geometry. In an image with all walls present, the technique can be outlined as follows:

- o The mitral valve plane appears as the flat side of the heart-shaped ANT view. It is normally located at the top left-hand side of the heart image, sloping leftward at about 45 degrees with respect to the image frame. In the LAO views, the valve plane is tangent to the gap in the ring. In the 45-LAO view the valve plan is approximately level with the image base; in the 67-LAO view, it slopes rightward at about 45 degrees.

- o The inferior cardiac wall can be found by locating the point on the interior of the ANT image opposite the valve plane, then drawing a line from that point to the outer wall sloping leftward at about 45 degrees. The inferior wall corresponds to the image region above this line.

- o The cardiac apex is the region between the above-mentioned line and a second line sloping 45 degrees rightward.

- o The anterolateral wall comprises the remaining image region in the ANT view.

- o The anterior, posterior, inferior, septal, posterolateral and inferioapical walls can be located by similar construction techniques on the two LAO views (reference figure 2).

The expert also examines wall contours. Wall edges should be approximately smooth. An indentation represents a perfusion defect at the edge of the wall (informally called a "rat-bite"). The indentation may extend along most of the wall's length, making the wall appear thin. The DRG feature extractor could detect these conditions by fitting an idealized image outline against the real image by adjusting the outline's position, orientation and scale. This is a form of template matching.

Perfusion defects inside a wall (informally, "bubbles") appear as enclosed regions whose mean pixel value is significantly lower than the neighborhood's. With appropriate pre-processing, the bubble's edges should be caught by the edge detector. The feature detector would then examine the number of standard deviations by which the average interior value differs from the average exterior, and decide whether or not to qualify the structure as a diagnostic feature. The qualifications should be liberal--the feature extractor must not make diagnostic decisions. It should pass on to the learning system any feature which a physician might conceivably use in making a diagnosis.

Identifying reperfusion defects requires comparing pixel intensities in the exercise and rest images. The naive approach would be to overlay the two images, subtract corresponding pixel values, and look for large differences. In practice, it is not possible to overlay the images. Considerable care is exercised to make the exercise and rest images agree in position, scale and orientation--a laser will shortly be installed on the camera to help replicate the patient's position. However, the ventricle itself constantly changes in shape due to normal systolic action. This prevents an accurate image overlay from being made.

The expert makes cross-image comparisons based on an informal visual mapping of corresponding image zones. The DRG bubble detector would divide each image into a gridwork of zones based on anatomical landmarks and geometry. A zone is intended to represent the same region of heart muscle in every image, even though zone size may change from image to image.

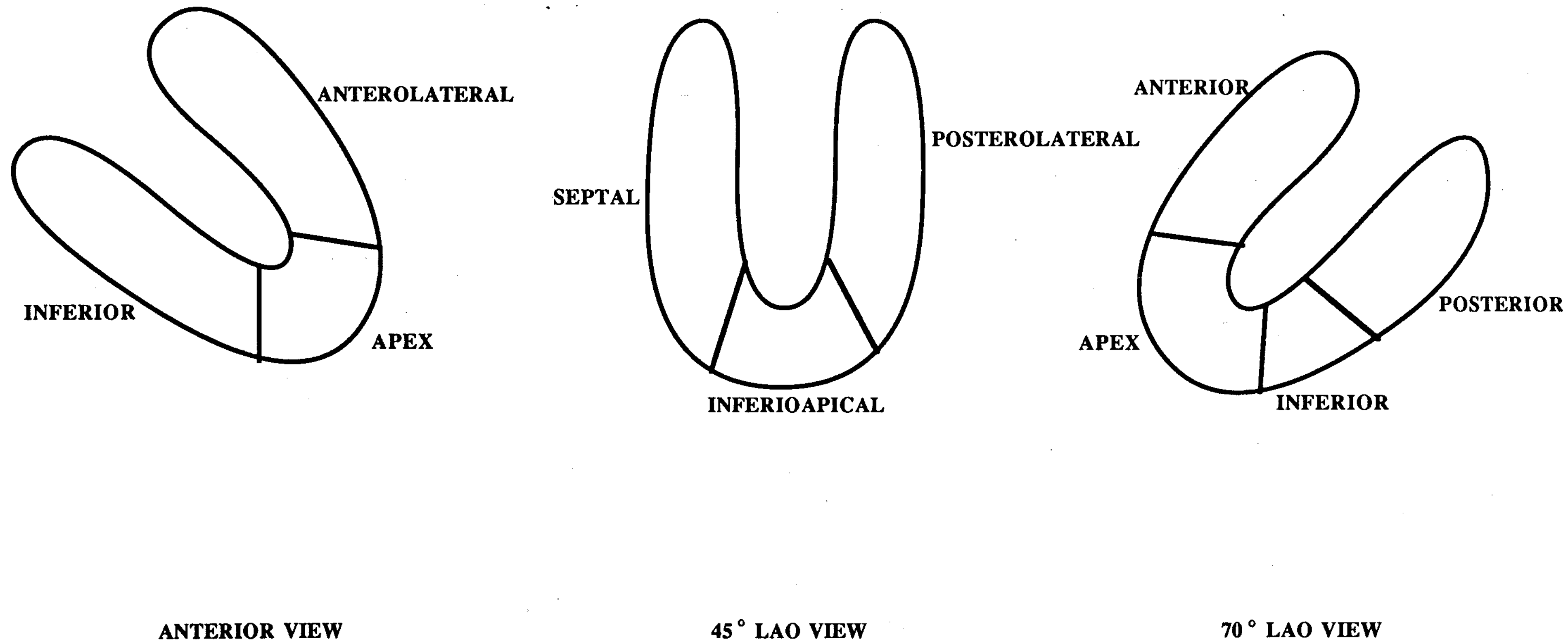

Figure 2. Left Ventricular Muscle Walls Appearing in AFSAM's Standard Views

The mean pixel value of zones can be compared between the exercise and rest images. This corresponds to what the expert does visually. (This zonal method is similar to the method used by Watson et al. in their automated thallium analysis system [WAT81]).

Feature extraction can be implemented using template matching techniques--in this case, curve fitting and geometric construction. These are conservative techniques with a long history in computer science. Neural nets could be used for some feature extraction tasks. Enjoying a recent renaissance, the technology dates back to the 1940s and is fairly well understood. The more complex and sophisticated techniques of syntactic analysis are not appropriate for the thallium domain, because the imagery is two-dimensional rather than a linear sequence of features in space or time, and because it is possible to identify features with little reference to other features. Performing feature extraction steps in the order discussed above should suffice to resolve what interdependence there is.

Feature extraction forms a bridge between the numeric and symbolic domains of telemetry analysis. Raw thallium data are numbers, and pre-processing manipulates them with standard mathematical techniques. But the feature extractor output will be symbolic ("a cold spot") as well as numeric ("over two standard deviations colder than the mean").

The feature extraction algorithms require elaboration and development. It will take time and effort to implement a feature extractor that works well in practice. However, the above discussion shows that there is little doubt that it can be done.

3.3 Risk Assessment for the Learning Subsystem

Generalizing rules from examples is not part of expert system technology, but the AI field of machine learning. This field is still very much a laboratory research area. (There are a number of commercial expert system shells which claim to operate on examples. None perform adequately on any but the simplest of problems. [TOM86.])

Do laboratory systems meet the requirements for machine learning about medical telemetry? To answer the question, the requirements must be identified.

A machine learning system for medical telemetry must handle several kinds of input data. Some data will be nominal, e.g. the presence of a "cold spot" in a thallium image. Other data will be numerical. In EKG interpretation, for example, a sinus rate over 100 means tachycardia. The learning system must be able to generalize about ranges of values. Finally, data may be hierarchically organized. The learning system must recognize that sinus tachycardia and supraventricular tachycardia are both tachycardias, so that it may learn about tachycardias in general.

The learning system must tolerate counterexamples in the data. Many learning algorithms will learn a rule only if the rule is never contradicted. Medicine is not that exact a science. Expert judgment will not be 100 percent consistent. A few anomalous cases should not always invalidate a rule.

The learning system should attach confidence factors to its rules. Since rules may not valid be for all the cases from which the system learns, the user must know how often a rule can be expected to be correct.

Some learning programs try to attain goals ("model-driven learning") while others let the data guide the process ("data- driven learning"). The latter strategy is unsuitable for inexact

problem domains; by beginning with anomalous data, the system may build an inductive castle on sand. A biomedical learning program should be model-driven.

The model will contain domain-specific knowledge to guide the induction process, e.g. knowledge about thallium imagery or EKGs. The model must be easily substitutable for other domain models. The system must not be specialized around one specific application.

All other things being equal, preference should be given to rules which are simple, readable, and make sense to human experts. The learning system should be biased towards simple rules.

The system should be capable of consulting the expert to guide the induction process at need. There are two ways to do that: explicitly and implicitly. In explicit consultation, the system asks questions of the expert. Implicit consultation means that the system produces alternative answers for the expert to evaluate and select.

A number of well-known induction algorithms were evaluated on these criteria (table 1). The evaluation matrix shows that none of these systems meets the requirements.

The feasibility of machine learning from medical telemetry is not demonstrated by existing systems. To evaluate the feasibility, an induction algorithm called METARULE was designed to meet the requirements, and prototyped as part of this study. METARULE is described in section 4 and evaluated in section 5.

3.4 Risk Assessment for the Rule Writer

The rule writer translates METARULE's internal rendering of diagnostic rules into a form acceptable to an external expert system shell. It may be desirable to have more than one rule writer in order to support more than one shell.

This is a straightforward task because METARULE's internal representation for rules is composed of the same elements used by most expert system shells: Boolean expressions involving tests on symbolic or numeric variables, and confidence factors.

How is the translation to be accomplished? METARULE's output is in a syntax known to logicians as conjunctive normal form. A METARULE result looks like:

```
IF
    (condition1 AND condition2 AND...) OR
    (condition3 AND condition4 AND...) OR ...
THEN
    the case is (NORMAL/BORDERLINE/ABNORMAL).
```

Where is certainty factor

In the antecedent expression, the first-level terms are connected by ORs. All lower-level terms are joined by ANDs. This is a syntax readily accepted by almost all expert system shells (and is the only syntax supported by the PROLOG language, which can be considered a shell of sorts).

When translating METARULE's conclusions, the rule writer need concern itself not with the rule's syntactic structure, which is nearly universal, but with semantic details. These include the proper words for IF and THEN, assigning legal variable names, using the proper notation for

	CATEGORICAL DATA	NUMERICAL DATA	HIERARCHICAL DATA	TOLERATES COUNTER EXAMPLES	ACCEPTS DOMAIN-SPECIFIC HEIRISTICS	DESIGNED FOR SPECIFIC DOMAIN	MODEL DRIVEN	EXPERT CAN GUIDE THE SYSTEM	ALTERNATIVE RULES	REFERENCE
C-NOTES	●		●							WINTS
ID3	●			●						QUI79
SPROUTER	●									HAY77
THOTH	●			●						VER80
INDUCE 1.2	●	●	●				●		●	MIC83
META-DENDRAL	●					chemistry	●		●	BUC78
CLUSTER/2										LEN8
BACON.4		●				chemistry	●	●	●	LAN81
AM	●	●				mathematics	●		●	LEN83
METARULE	●	●	●	●	1		●	●	●	THIS REPORT

Table 1. Comparison of METARULE with
Several Well-Known Machine Learning Systems

test operations (A EQUALS B versus EQUAL(A,B) versus A = B versus A IS B), how to write a confidence factor, etc. The only syntactic translation expected to be required is when the order of the antecedent and consequent are inverted (IF A THEN B versus B IF A); and the proper use of parentheses.

It would be possible to write an elegant rule- (or grammar-) based translation program that produces translated output as a side-effect of parsing input from METARULE. Given the relatively simple nature of the translation involved, it would be just as effective (and much faster) to write a translation program using conventional programming techniques. The latter approach is recommended.

METARULE already contains a variant on the rule writer. The METARULE component that produces structured English descriptions is a rule writing module intended to produce readable output. Producing human-readable rules is a more difficult task than producing machine-readable rules. Although the structured English translations produced by the METARULE prototype could be improved, the programming technology used (parsing by recursive descent) is more sophisticated than will be necessary for most expert system shells.

The DRG rule writing subsystem poses no significant technical risks. impossible

Example

If A and B then C.

In the above example, what is the certainty of C?

a. The min certainty of A, B

b. The product of the certainties of A [illegible]

c. Other, e.g. Bayesian rule of probability [illegible]

If A or B then C [illegible]

In the above example, what is the certainty

a. The max certainty of A, B

b. The min certainty of 1, A+B

c. Certainty based on [illegible] logistic risk equation (Walker & Duncan) with main effects and interactions

4.0 METARULE: A PROTOTYPE MACHINE LEARNING SYSTEM FOR DRG

To explore the feasibility of machine learning under the demands place by thallium imagery, a new machine learning system was designed and prototyped. The system is called METARULE because it learns "about rules" from examples.

METARULE was tested on a variety of training sets. Several of these were abstract problems specially constructed to stress METARULE's capabilities. In addition, METARULE was run on a set of feature data derived from thallium information supplied by USAFSAM. The results of these experiments are evaluated in section 5.

METARULE was prototyped in the INTERLISP-D language on a Xerox 1186 LISP machine (about 2700 lines of code). The LISP language was chosen because experience with LISP versus conventional ("C") language development at Analytics has shown that complex systems can be prototyped 3 to 10 times faster in LISP. The prototype system is subsequently ported to C for delivery. For METARULE, a conscious effort was made to avoid LISP features that are difficult to replicate in other languages, e.g. no routines that rewrite themselves while running.

This section first explains the method of knowledge representation which METARULE uses internally for formulating and testing hypotheses. Next, the METARULE learning algorithm is described. A final subsection covers the structured English rule writer.

4.1 A Language For Describing Hierarchically Organized Case Data

The choice of a method for representing knowledge is fundamental to the success of an artificial intelligence system [BRA85]. Examples of representations used in AI include rules, frames, first-order logic, object-instance hierarchies and procedures.

The method of representation (often called the representational scheme) forms a kind of language in which the computer reasons about the problem at hand. The language must be appropriate to the problem. Just as some problems are easier to solve when posed as mathematical equations, while others are better expressed in words, so the representational scheme used by an AI system should follow the natural contours of the problem.

Two desiderata for a representational scheme are expressiveness and rigor. Expressiveness means that the scheme should be able to represent as much of the wealth and subtlety of the real world as is required to manifest (simulated) understanding. A non-expressive scheme is a Procrustean bed into which knowledge and information must be force-fit. Rigor means that the scheme is well-defined: every permitted representation has a definite meaning. That is not to say that uncertainty has no role in knowledge representation. If uncertainty is to represented, it should be explicitly represented in a clear and definite way. There should be no uncertainty, however, about what a statement in the representational language means.

There are a number of styles for representational schemes. These include declarative representations, of which rules are the most familiar example. Declarative procedures are knowledge about "what" as opposed to knowledge about "how". In a traditional expert system, the procedural knowledge about "how" to reason using rules is called the "inference engine" and is segregated from the declarative rule base. AI systems based on first-order logic, such as those written in the PROLOG language, are also declarative. MYCIN, the original expert system for the diagnosis of infectious diseases, exemplifies a medical application employing declarative representation.

Procedural knowledge is knowledge about "how to". Heuristics (an expert's rules of thumb) underlie many procedure-oriented AI systems. Unlike declarative representations, which simply make a series of statements about the world, procedural representations encode expertise as a series of steps to be followed.

There are other styles as well such as probabilistic modeling, connectionist models (simulated neural networks which have adaptive, self-organizing properties), qualitative models (which include notions of causality and time). Causal modeling is employed by ABEL, an expert system that diagnoses electrolyte disturbances [PAT82]. ABEL knows, for example, that metabolic acidosis causes acidemia, which attenuates hypokalemia and may cause hyperventilation. Such causal knowledge helps ABEL reason in an efficient and realistic manner.

For learning diagnostic criteria for thallium imagery, METARULE employs a representational scheme in the declarative style that uses both semantic nets and first-order logic. Since a goal of representation is to follow the natural contours of the knowledge to be represented, the scheme can best be explained and justified by examining the structure of the problem.

4.1.1 Data Received from the Feature Extractor

The data from which METARULE formulates its diagnostic criteria consist of a set of images paired with diagnoses. The purpose of the set of image/diagnosis pairs is to train the system how to perform diagnosis using the criteria implicit in the set. Such sets will be called training sets. Each image/diagnosis pair will be called a case.

The diagnosis of a case is a nominal variable which may have one of three values: "normal", "abnormal" or "borderline".

The "image" in a case is a list of features identified by the feature extractor. In the case of thallium, these are expected to include:

- o missing walls;
- o localized perfusion defects, occurring in exercise images;
- o localized reperfusion defects, occurring in rest images; and,
- o matched defects, occurring in both images.

Each feature will have attributes with corresponding values. A preliminary list of attribute is:

- o the view(s) in which the defect occurs;
- o the wall(s) in which the defect occurs;
- o the portion of the wall in which in which the defect occurs (this may be significant for defects near the valve plane or the apex);
- o the defect type ("bubble", "rat-bite", etc.);
- o the percentage of wall thickness involved;
- o the percentage of all area involved;

- the intensity of the defect, measured perhaps as the number of standard deviations by which the mean pixel value within the defect varies from the mean pixel value of the ventricle;

- for matched defects, the degree of washout.

Other feature types and/or attributes may appear in the fully developed DRG.

METARULE allows features to belong to classes. In some medical domains, certain features may suggest the absence of disease while others suggest its presence. Such features do not seem to occur in thallium--the only features mentioned by the interviewees are pathological.

In the METARULE prototype, a case is represented as a LISP list having the structure:

```
(DIAGNOSIS   (FEATURE-TYPE  FEATURE-ID (ATTRIBUTE VALUE)...)
             (FEATURE-TYPE  FEATURE-ID (ATTRIBUTE VALUE)...)
e.g.,
             (BORDERLINE (PERFUSION-DEFECT P1 (INTENSITY 2)
             (LOCATION INFERIOR-WALL) ....)
             (MATCHED-DEFECT M1 (INTENSITY .5)
             (LOCATION SEPTAL-WALL)...))
```

Each case has a name. A training set is a list of case names:

```
(CASE1 CASE2 CASE3 ...)
```

Feature classes are treated differently, because they are not derived directly from image data, but from background knowledge which the expert brings to the problem. Therefore feature classes are not output by the feature extractor, but are part of the thallium-specific knowledge base in METARULE.

4.1.2 METARULE'S INTERNAL REPRESENTATION OF A TRAINING SET

The expert's view of thallium features is hierarchically organized. A feature belongs to a case and also to a class. A feature has one or more attributes, which have values. Since a feature belongs to two distinct entities (a case and a class), the hierarchy is of a mathematical type known as a lattice (as opposed to a simpler kind of hierarchy, a tree).

The relationships among these entities can be expressed using a representational scheme known as a semantic net [QUI68] (figure 3). Training sets produced by the feature extractor are organized using the semantic net as a template. METARULE's learning module uses the semantic net to interpret case data, and understand the relationships among the entities that make up a case.

The semantic net suffices to describe a case. But to learn diagnostic criteria, something more is needed. METARULE's learning module formulates, tests and refines hypotheses about what makes a case abnormal (or normal, or borderline). The representational scheme must be able to state, evaluate and manipulate such hypotheses.

To that end, METARULE's semantic net is augmented with a system of predicates. A predicate is a formalism used in first- order logic to make assertions about the world [AND86]. For example, the first-order expression "EXPERT(JOHN)" asserts that John is an expert. "EXPERT" is an example of a predicate. A predicate expresses a quality which is true of its argument. A predicate expression is either true or false when applied to a specific case--in the above example, John may or may not be an expert. An assertion involving a predicate will be called a "clause".

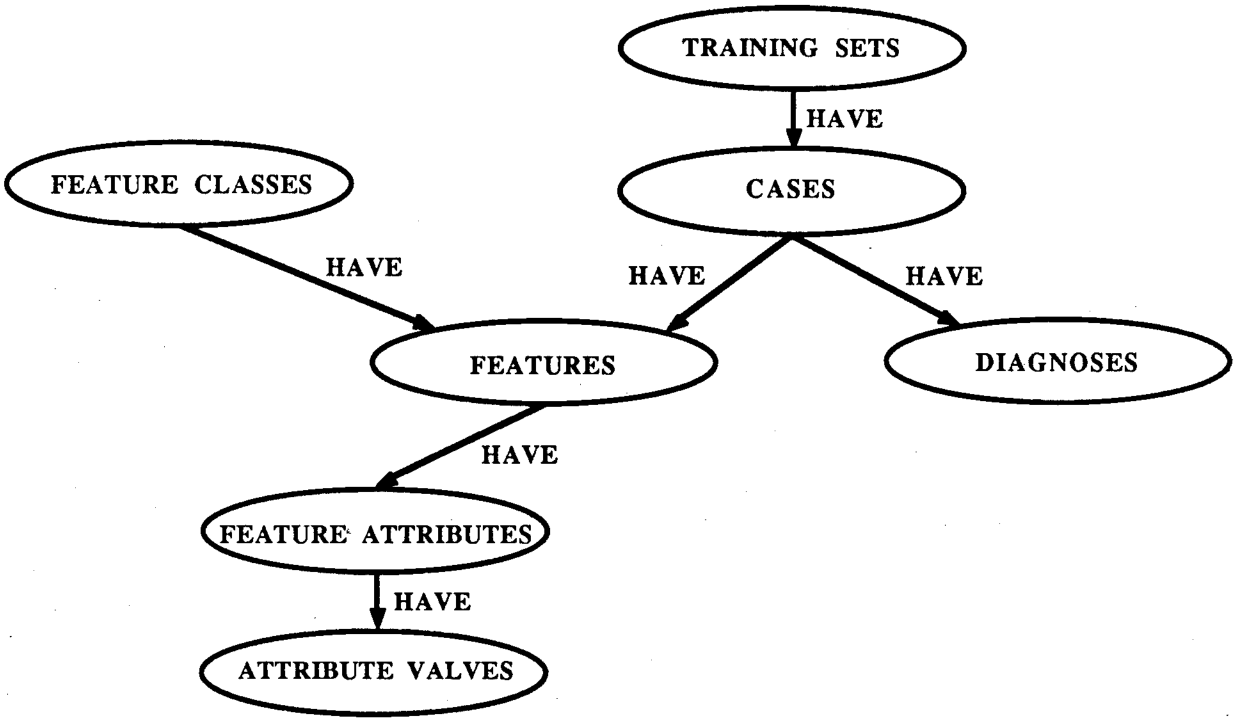

Figure 3. The Semantic Net

The argument of a predicate may be a literal value as in EXPERT(JOHN), or another predicate, as in GREATER- THAN(AGE(JOHN),10).

METARULE uses fourteen predicates to formulate diagnostic criteria. Three predicates identify the features, feature types, attributes and values present in a case:

- o HAS.TYPE(X) asserts that the case contains at least one feature of type X. A feature type is a "perfusion defect", "reperfusion defect", etc. The feature type is a property of an individual feature, which is identified by a serial number. There may be more than one feature of a given type in a case.

- o HAS.ATTRIB(X,Y) asserts that the case contains at least one attribute X whose value passes test Y. For example, "HAS.ATTRIBUTE(WASHOUT.RATE,(GT VALUE .5))"

- o HAS.CLASS(X) asserts that the case contains at least one feature whose type belongs to class X.

Three additional predicates are used for constructing tests on attribute values:

- o EQUAL(X,Y) asserts that X is equal to Y. X and Y may be numbers or categorical values like "normal", "high", etc.

- o LT(X,Y) asserts that X is less than Y. Both X and Y must be numbers.

- o GT(X,Y) asserts that X is greater than Y. Both X and Y must be numbers.

Four predicates are used to locate features, attributes and classes on the semantic net. Each clause generated by METARULE is identified by a unique "clause number". The next four predicates take clause numbers as their arguments.

- o TRUE.OF.SAME.CASE(X,Y,Z,...) asserts that the clauses identified by clause numbers X, Y, Z and so forth are all true of the same case. This predicate performs the function of the Boolean AND operator.

- o TRUE.OF.SAME.FEATURE(X,Y,Z...) asserts that the specified clauses are all true of the same feature.

- o TRUE.OF.SAME.TYPE(X,Y,Z...) asserts that the specified clauses can be satisfied by a set of features of the same type.

- o TRUE.OF.SAME.CLASS(X,Y,Z...) asserts that the specified clauses can be satisfied by a set of features of the same class.

Two additional predicates are used for counting:

- o NUMBER.OF.FEATURES(X) asserts that the case has X features.

- o NUMER.OF.FEATURES.MEETING.CLAUSE(X,Y) asserts that clause X is true of Y features in the case.

The final two predicates provide the remaining basic logical operations:

- o NOT:(X) asserts that clause number X is false.
- o OR:(X,Y,,,) asserts that at least one of its argument clauses is true.

It should be noted any structure describable by the semantic net of figure 3 can be described using the above predicates. In turn, any structure describable by the semantic net is also describable by the syntax specified for the feature extractor's output. This demonstrates the expressiveness of all three representations with respect to the expert's view of the image.

To make the representation rigorous, a few semantic rules must be adduced:

- o The TRUE.OF.SAME.CASE predicate can take any types of clauses among its arguments;
- o The TRUE.OF.SAME.FEATURE predicate can take only clauses whose predicate is HAS.FEATURE, HAS.ATTRIBUTE or NUMBER OF FEATURES MEETING CLAUSE among its arguments, and clauses whose predicate is NOT: which negate one of the above-mentioned clause types;
- o The TRUE.OF.SAME.TYPE and TRUE.OF.SAME.CLASS predicates can only take clauses whose predicate is HAS.ATTRIBUTE and NUMBER OF FEATURES MEETING CLAUSE, and clauses whose predicate is NOT: which negate one of the above-mentioned clause types.

These rules prevent METARULE from writing clauses that violate the hierarchy specified by the semantic net. Such as clause would violate the structure of an expert view of the data, e.g. having a case that belongs to a feature instead of vice versa.

An additional rule prevents METARULE from constructing double negatives:

- o The NOT: predicate may be applied only to the HAS.TYPE, HAS.ATTRIBUTE and HAS.CLASS predicates.

It should be reemphasized that the semantic net and the predicate logic are used internally by METARULE. These representations are designed for knowledge representation and reasoning, not for display to the user.

4.1.3 METARULE'S DOMAIN MODEL

METARULE incorporates a domain model for thallium. The model contains expert knowledge that enriches and guides the induction process. The model is expressed by representations which are separate from the induction algorithm. Thus, METARULE is not specialized around a particular application. The model is contained in a single LISP function which can be easily substituted to tell METARULE about other domains.

A METARULE domain model contains three kinds of knowledge.

The first kind of knowledge is about derived features and attributes. These are features and attributes which do not directly appear on the image, but whose presence can be deduced from other image properties.

In thallium interpretation, the distribution of features among segment walls is an important diagnostic clue. Defects which lie within the distribution of a coronary artery strongly

suggest the presence of CAD in that artery. Also, the diagnostic significance of a defect depends on the defect's location. An apparent defect observed near the mitral valve plane is not a reliable indicator of CAD.

The METARULE thallium domain model contains rules for adding derived features and attributes:

- o A derived feature of LAO-CAD-DISTRIBUTION is added if there is a defect in at least two of the following wall segments: apical, septal, anterior.

- o A derived feature of CIRC-CAD-DISTRIBUTION is added if there is a defect in at least two of the following wall segments: apical, posterolateral, posterior, anterolateral, inferior.

- o An attribute called RELIABILITY is added with a value of LOW to a feature located near the valve plane, or with a value of HIGH if the feature occurs elsewhere.

A second kind of domain knowledge provides the induction algorithm with generally accepted background information to guide the search for rules. Background information is provided as a set of features or attribute values which are associated with certain findings:

- o Normal findings are associated with:

  - The absence of abnormal features.

- o Borderline findings are associated with:

  - Reperfusion defects;
  - High reliabilities;
  - Small wall thicknesses; and,
  - Small numbers of wall segments involved.

- o Abnormal features are associated with:

  - Matched defects;
  - Reperfusion defects;
  - High reliabilities;
  - Small wall thicknesses; and,
  - Large numbers of wall segments involved.

The third kind of domain knowledge sets parameters governing the induction process, e.g. the criterion for qualifying an hypothesis as a rule, the number of hypotheses to explore, etc. These parameters are discussed more fully in section 4.2.

METARULE's domain knowledge is represented as LISP data structures. For example, the background knowledge concerning normal cases reads in LISP: "(NORMAL (CLASS NOT ABNORMAL))". It would be desirable to have a user interface by which the expert might enter and modify such knowledge; however, a user interface to the domain model was not built for the prototype.

4.2 METARULE'S INDUCTION ALGORITHM

Sophisticated induction algorithms are based on intelligent search strategies. An algorithm seeks solutions--generalizations that classify all or most of the cases correctly. Every generalization that can be stated in the representational scheme is a possible solution. This is a vast number of generalizations; not all of them can be explored in a reasonable amount of time.

An induction algorithm must therefore do two things. First, it must limit the number of generalizations that are explored (or equivalently, limit the time spent exploring generalizations). Second, it must employ an intelligent strategy to search out solutions early on. There is no guarantee that the best solution will be found in the allotted time. That is true of sophisticated induction algorithms in general. Even so, good solutions may be found. The quality of the solutions will depend on the cleverness of the search strategy and the length of the search.

The goodness of a solution depends on how many cases it diagnoses correctly. A good solution minimizes two kinds of incorrect diagnoses: false positives and false negatives. Unless an infallible diagnostic criterion exists, there is a tradeoff: diagnostic criteria designed to minimize false negatives will raise the probability of a false positive and vice versa.

The user may value one kind of goodness over the other. USAFSAM cares more about reducing false negatives than reducing false positives; false positives will be eliminated by subsequent angiography.

METARULE enables the user to specify the desired tradeoff. The goodness of a solution is quantified as the weighted sum of the fraction of correctly diagnosed positive and the fraction of correctly diagnosed negative cases. The user can set the weight to any value between 0 (eliminating false positives is the only goal) and 1 (eliminating false negatives is the only goal).

Simple rules are easier to understand, validate and apply. METARULE incorporates an inductive bias towards simplicity in its solutions [UTG86]. The bias is achieved in several ways:

- Simple candidate solutions (hypotheses) are tried first.
- Complex hypotheses are reduced to their simplest logical equivalent.
- Solutions which have simpler versions that perform at least as well are rejected in favor of the simpler version.
- If two solutions that are not logically equivalent diagnose all the cases the same way (functional equivalence), the user is asked if the two are equivalent in general. If so, the user will be asked which hypothesis to retain.

This winnowing process has the beneficial side-effect of speeding the search. METARULE searches by building more complex hypotheses from promising simpler hypotheses. For each eliminated hypothesis, METARULE saves the time it would otherwise spend trying to elaborate the hypothesis in various ways.

METARULE also winnows the search space is by selecting only the most promising hypotheses for elaboration. Each time a new hypothesis is generated, it is evaluated. Only the N most promising hypotheses are candidates for elaboration. For the runs in this report, N=50. Thus, if there are 50 candidates for elaboration and a new hypothesis performs better than the weakest candidate, the new hypothesis will be added to the list of candidates and the weakest one removed.

METARULE begins its search by hypothesizing that some single feature type, attribute value, or feature class is what makes a positive case positive. Then it hypotheses about the combinations that appear in the training set. For example, if a positive case contained a perfusion defect with intensity 2, METARULE would postulate that:

- a reperfusion defect makes a case positive;
- a feature with intensity 2 makes a case positive; and,
- a reperfusion defect with intensity 2 makes a case positive.

For negative cases, METARULE postulates that each feature, attribute value, class and combination prevent the case from being positive.

METARULE performs special processing on attributes with numerical values. Preliminary hypothesis generation as described above will hypothesize that the diagnosis depends on a numerical attribute having specific values. But the diagnosis may depend on a value falling within a given range. METARULE generalizes hypotheses about numerical values into hypotheses about ranges.

By way of illustration, reperfusion defects with intensities greater than 2 standard deviations are associated with a diagnosis of abnormal. In a training set, a range of actual intensity measurements will occur. METARULE will generate a "cutpoint hypothesis" that a positive case must have a value greater than (or less than) a certain value (the cutpoint).

Cutpoints are generated by listing all the numerical values that an attribute assumes, then identifying the value with the greatest discriminatory power.

METARULE's domain model assists in selecting cutpoints. There may be more than one powerful cutpoint. That is especially true for small training sets. If so, METARULE consults the domain knowledge. The domain model may specify that high or low values of the attribute are associated with the diagnosis of interest. If so, METARULE will select the highest (or lowest) cutpoint with the maximal discriminatory power.

One or more solutions may result from preliminary hypothesis generation. A solution is an hypothesis whose goodness rating exceed a user-specified threshold. For the runs in this report, the threshold was set to .5 . The choice of a low threshold has a purpose: since METARULE retains only the best N=50 solutions, a process of natural selection will crowd out the weaker solutions if better ones are obtained. If better solutions do not emerge, solutions at the .5 level will be reported.

Next, METARULE searches for better solutions by combining hypotheses. Recall that METARULE solutions are in conjunctive normal form. The first step generates disjunctive clauses by grouping existing clauses together. This is done by the use of the grouping predicate TRUE.OF.SAME.CASE, which is equivalent in function to the Boolean "AND" operator.

How does METARULE decide which combinations to try? METARULE has a basic strategy with a number of elaborations.

An hypothesis can be regarded as either of two kinds of generalization: a characteristic description or a discriminant description [MIC83]. A characteristic description lists all the things that positive cases have in common; a discriminant description tells what separates positive cases from negative cases. Solution clauses are discriminant descriptions.

If an hypothesis is not a good discriminant description, it may be useful as a characteristic description. A characteristic description covers the positive cases well, but include too many negative ones. In a sense, a characteristic description is half a solution (few false negatives) in search of its other half (an hypothesis that reduces the number of false positives).

METARULE begins its search with the strongest characteristic description, exploring descriptions in decreasing order of strength. The N strongest characteristic descriptions are examined. (For the runs reported in section 5, N=30.)

For each characteristic description, a set of promising hypotheses for disjunction are identified. Disjunction reduces the number of positive cases covered by a description. A good disjunctive clause will reject false positives allowed by the characteristic description while preserving the true positives.

METARULE generates a list of promising disjunctive hypotheses for each characterizer. These are ranked from strongest to weakest. In that order, METARULE forms new hypotheses by disjunction. Only the best N percent of the disjunctive clauses are explored (for the runs in section 5, N=25%.)

As new hypotheses emerge, the list of characterizers is constantly updated. If the disjunctive clause is a promising characterizer, it may be consecutively elaborated several times.

This strategy implements a type of search technically known as a beam search. The search space is the tree formed by the disjunction of all possible hypotheses in all possible combinations. METARULE does not generate this enormous tree completely, but starting from a promising position, explores the branches that appear most promising. At each step, branches may be added or pruned from the active set. The set forms a kind of "beam" illuminating the most promising parts of the tree; the rest remains in darkness, unexplored.

There are a number of elaborations on this basic strategy.

Some are intended to narrow the search space.

A clause which covers all cases is useless as a disjunctive clause (since the disjunction will make the same predications as the characteristic clause). Such clauses are not used in disjunction.

Also, the search strategy may uncover disjunctions which are logically (or empirically) equivalent to other disjunctions. These are rejected during the search.

The logical equivalence of two clauses is determined by demonstrating that one clause can be simplified to produce the other. The semantic net states inclusion relationships among predicates that can be used to simplify disjunctive clauses. The rules that METARULE uses for simplifying disjunctive clauses are:

- If a disjunction references a NUMBER.OF.FEATURES.MEETING.CLAUSE predicate or a NOT:(NUMBER.OF.FEATURES.MEETING.CLAUSE) compound predicate with argument X, and clause X elsewhere occurs in the disjunction, then remove the reference to clause X. (For example, If the hypothesis is that there are two perfusion defects in an abnormal case, it is redundant to say that there must also be "a" perfusion defect).

- o If a disjunction contains a grouping predicate which has clause X among its arguments, and X appears separately in the disjunction, then remove the reference to clause X.

- o If a disjunction contains the same grouping predicate twice, and one predicate's arguments are a subset of the others, then remove the predicate with the smaller set of arguments.

Two clauses are empirically equivalent if they perform identically on the training set. This might be accidental, an artifact of the cases selected of training. Or, the equivalence might reflect some deeper knowledge about the domain. For example, if there were only one feature (e.g. FEVER) with a class of "abnormal" in some domain, METARULE would detect the empirical equivalence of the clauses

(HAS.CLASS ABNORMAL)

and

(HAS.FEATURE FEVER).

METARULE would then ask the user if the clauses are equivalent in general and, if so, which one the user prefers. The non-preferred clause will not be used in the search process. However, it will be reported among the solutions.

Other elaborations are intended to promote a bias towards simple rules. Characteristic descriptions and disjunctive clauses are sorted by their descriptive power. For clauses whose powers are equal, preference is given to simpler clauses. These are clauses which have a fewer number of components. Clauses generated earlier in the search process are likely to be simpler; therefore, all other things being equal, older clauses are given preference.

A third kind of elaboration makes the search smarter, i.e. gives precedence to hypotheses that are likely to be the seeds about which a solution crystallizes. This is accomplished via a mechanism called "bias". Bias is a property that may be bestowed upon a clause. Having bias gives the clause precedence in the search process.

One way that an hypothesis can acquire bias is by making predictions about numerical ranges. Bias is automatically bestowed on these hypotheses to boost them ahead of hypotheses about specific numerical values.

METARULE's domain model can also give bias to an hypothesis. For example, the thallium domain model specifies that abnormal cases are associated with matched defects. If the diagnosis of interest is "abnormal", METARULE will grant bias to any hypothesis that specifies the presence of a matched defect.

Finally, METARULE seeks conjunctive solutions. It identifies solutions which are strong but not perfect in their coverage of positive cases, and admit few false negatives. These are sorted by their strength as characterizers. A beam search is performed to identify complementary clauses that correct each other's false negatives while admitting few false positives. If found, the two characterizers are conjuncted ("OR'd").

Most expert system shells associate certainty factors with their rules. For shells based on Bayes' law, the certainty factors are probabilities. Other shells use a MYCIN-like calculus for propagating certainty [BUC84].

METARULE attaches a certainty factor to its solutions. The certainty factor is the probability that a case meeting the solution clause has whatever attribute has been designated as the

Certainty Factor

"positive" attribute. The probability is, of course, a frequentist probability based on METARULE's experience with the training set.

In outline, the METARULE induction algorithm is:

1). Read the training examples and represent them internally as an instantiated semantic net. Instantiated means that an actual value is assigned to a field or attribute of a fact

2). Generate an initial set of hypotheses (candidate discriminant descriptions). The initial descriptions are simple, giving METARULE a bias towards simplicity.

3). As each hypothesis is generated, test its predictive power on the training set and assign it a numerical score. If the score exceeds the user-specified solution threshold, place the hypothesis on the list of solutions. The hypothesis is also placed on the list of characteristic descriptions.

4). Provide an additional bias toward simplicity by rejecting from the solution list any hypothesis which has a simpler equivalent that scores at least as well. Clause A is considered a simpler equivalent of clause B if the clauses to which B refers (directly or indirectly) are a subset of the clauses to which A refers.

5). When an hypothesis is placed on its list (characterizer or solution list), maintain the list in sorted order giving precedence to hypothesis with bias, high-scoring hypotheses, and simple hypotheses in that order. If the length of a list exceeds a preset maximum, trim the list by discarding the lowest-scoring hypothesis.

6). Generate additional initial hypotheses by applying domain-specific (in this case, thallium-imagery specific) background knowledge to the semantic net. Process each hypothesis as specified in steps 3-5.

7). Generalize numerical attribute values. For each numerical attribute, seek a cutpoint which divides the positive cases from the negative cases with an error rate less than the user's solution cutoff value. If a cutpoint is found, generate a hypothesis that a positive case must have a value greater than (or less than) the cutpoint.

8). Select the highest scoring characterizer which has not yet been explored. Construct a candidate discriminant list of clauses which reject the false positive cases admitted by the characterizer. The list is constructed in accordance with the procedure given in steps 3-4.

9). For the most promising clauses on the discriminant list, form a new hypothesis by conjuncting ("anding") it with the clause from the characterizer list.

10). Simplify the new clause if possible by combining or eliminating redundant predicates.

11). Check to see if the simplified clause is logically equivalent to an hypothesis already evaluated. If so, return to step 8.

12). Check to see if the simplified clause is functionally equivalent to another hypothesis, i.e. classifies each case in the same way. If so, call this to the attention of the user and ask if the hypotheses are equivalent in general, or if their equivalence is simply an artifact of the training set. If the expert says that the functional equivalence is true in general, ask which hypothesis the user prefers to

retain. The user will choose the hypothesis that seems simpler, more understandable, or more promising. The other hypothesis will be retained but not explored any further.

13). Evaluate the new hypothesis in accordance with steps 3- 5.

14). If the allotted number of hypotheses have not yet been explored, return to step 8.

15). Try improving the hypotheses on the solution list by disjuncting ("ORing") them in with other hypotheses Perform a beam search similar to that specified in steps 8-14.

4.2.1 The Structured English Translator

The METARULE prototype has a translator that renders clauses into structured English. When a run is complete, the user examines the rules in English translation. Rules are ranked in order of goodness, from the strongest to the weakest. The user views a rule by typing: RULE N, where N stands for the rank number of the rule.

In DRG, the user will select for use one or more rules, based on their goodness, clarity, generality, etc. The selected solutions are passed to the rule writer.

The translator prints rules in conjunctive normal form, using indentation to indicate the precedence of operations, e.g.:

A case is NORMAL if:

There are no features

OR

There is exactly one abnormal features

AND

There is an abnormal feature in the valve plane

Each term in the conjunctive expression is a sentence consisting of a verb phrase, number phrase, adjective phrase, noun phrase and a second adjective phrase in gerund-like form. For example, the clauses:

1 TRUE.OF.SAME.CASE(2,3,4)
2 COUNT.FEATURES.MEETING.CLAUSE (3, EQUAL(VALUE,1))
3 TRUE.OF.SAME.FEATURE(4,5,6,7)
4 HAS.CLASS(ABNORMAL)
5 HAS.LENGTH(2)
6 HAS.ATTRIBUTE(RED)
7 HAS.ATTRIBUTE (BIG)

would create the following phrases:

Verb phrase:	There is
Number phrase:	exactly 1
Adjective phrase 1:	big red
Noun phrase:	abnormal feature
Adjective phrase 2:	having a length of 3.

Which yield an English gloss of:

There is exactly 1 big red abnormal feature having a length of 3.

Each attribute known to METARULE has an attached property telling how to describe it in English. In the above example, the attribute RED has an attached description of ADJECTIVE, and LENGTH has a description of GERUND with an associated phrase "having a length of ?". When the description is transformed into a phrase, the value is substituted for the question mark.

The rule is followed by a short evaluation giving a confidence factor, e.g.:

This rule correctly classifies 83.3% of the NORMAL cases and rejects 100% of the other cases. The performance rating of this rule is .917

A difference between METARULE's structured English gloss and a conventional knowledge base is that METARULE reports a solution as one big rule. A knowledge engineer might have broken the solution down into several rules. This is more a matter of style rather than substance.

The rule writer should decompose a complicated rule into parts. This could be done by splitting the rule at the conjunctive level, e.g.:

if (A and B and C) or (D and E) then F

becomes:

if Y and Z then F
if (A and B and C) then Y
if (D and E and F) then Z.

5.0 EVALUATION

The feasibility of DRG has been shown to depend on the feasibility of machine learning. This section evaluates the METARULE prototype. Two kinds of evaluation are performed. First, METARULE capabilities are compared with the formal requirements developed in section 3.3. Then the operation of METARULE is described and evaluated.

METARULE satisfies all of the formal requirements identified for the medical telemetry domain:

- o generalization about nominal attributes;
- o generalization about numerical attributes (ranges);
- o representation of hierararchically structured data;
- o incorporation of a domain model;
- o a bias towards simple rules;
- o explicit and implicit expert guidance;
- o toleration of counterexamples;
- o reporting of certainty factors.

The specifics of how METARULE meets each requirement can be found in section 4.

Operational testing of METARULE on thallium data was limited by the fact that Analytics received suitable data less than two weeks before the final report's due date. In the interim, METARULE was tested against a set of artificial problems designed to test various capabilities.

The artificial problem sets concern a notional problem often used in the induction literature. It will be called the "Martian cell" problem.

Imagine that we study cells from a Martian organism whose biology is dissimilar to anything seen on Earth. These cells have walls and a cytoplasm with bodies of various sorts. The bodies have three properties: a shape, a color and a weight. Specifically, some of the bodies are circular and others are coiled forms resembling springs. Circles are considered normal; springs are abnormal features. The bodies may be red, white, green or blue. And some of the bodies weigh 1 unit while others weigh 2.

When cultured, certain cell types appear to be cancerous, in the sense of uncontrolled growth. METARULE is tasked with the following problem:

> From a training set of cell types together with diagnoses (cancerous or normal), discover what makes a cell cancerous.

Figure 4 shows a training set for this problem. The reader is invited to try solving the problem before proceeding. There is a rule which diagnoses all of the cases correctly.

METARULE was set to work on this problem with no domain knowledge, a weighting factor of .9 (strong bias towards avoiding false negatives), a elaboration limit of 30 (try disjuncting or conjuncting at most 30 characteristic descriptions), a branching factor of .25 (try disjuncting or conjuncting with the top-rated 25% of the candidate list), maximum candidate list length of 50, and a solution cutoff of .5

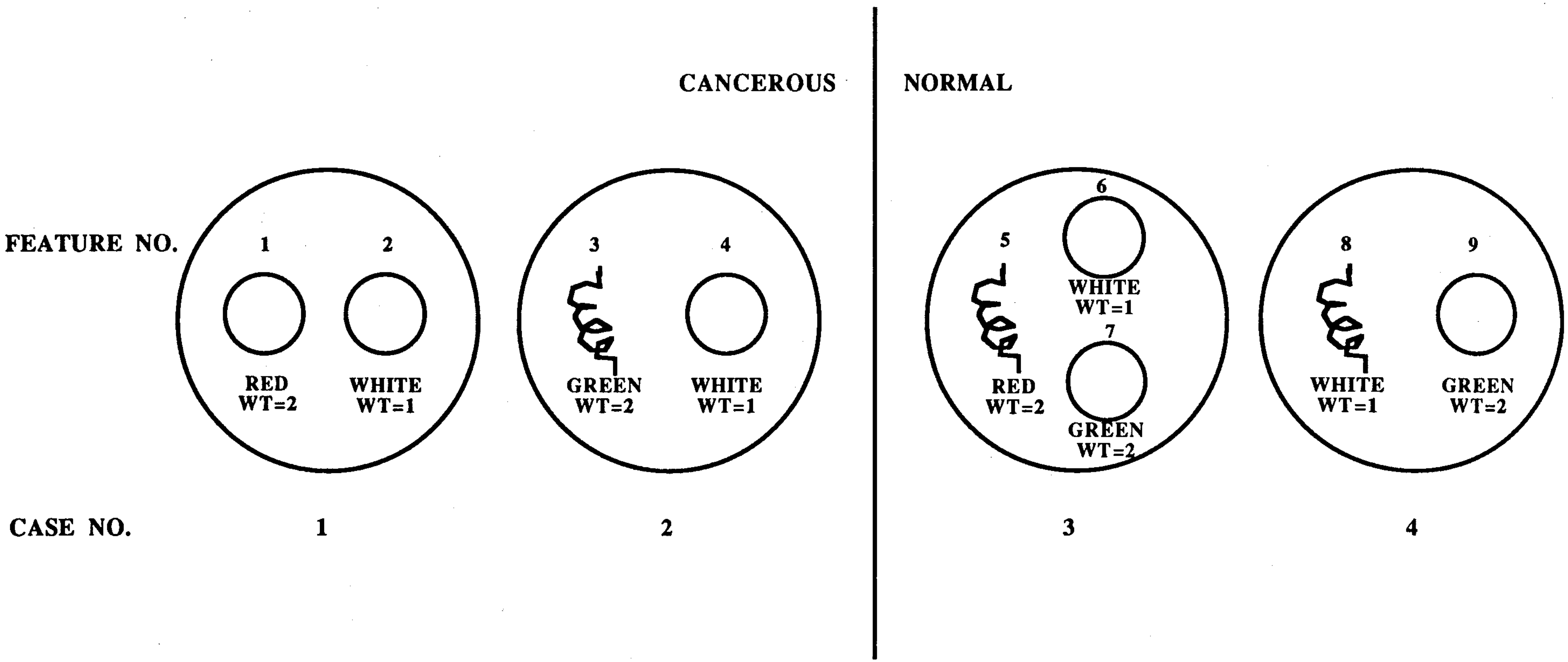

Figure 4. Test Problem Number One:
What Makes a "Cell" Cancerous?

The system ran for 24 minutes, generating 50 candidate rules. The top-ranking rule was:

A case is POSITIVE if:
there are circles such that their combined characteristics include the following:
white, weighing 1
AND
there are exactly two features in the case.

This rule does in fact solve the problem and is perhaps the most convincing rule to do so. An alternative rule is:

A case is POSITIVE if:
there are circles such that their combined characteristics include the following:
white, weighing 1
AND
the case has exactly one feature weighing 2.

METARULE found 48 additional rules, all of which diagnose the training cases correctly. The complete set of rules appears in appendix A.

Two comments need to be made about these results.

First, the structured English phrase "there are circles..." seems somewhat awkward and unnatural. Considerable thought was taken on how to word this phrase.

The problem is that certain METARULE predicates express abstract relationships which are inherently awkward to say in words. The above example is a gloss for TRUE.OF.SAME.CASE(...). In other words, there must be at least one feature in the case which is a circle, and at least one of the circles must be white, and at least one of the circles (but not necessarily the white one) must weigh 1.

If the prototype is developed into a full DRG, better phrasings should be sought for abstract METARULE clauses. It would be best to do this in consultation with future users.

Second, the run time of 24 minutes raises the question: does this time scale linearly in training set size? Would 100 cases take 2,400 minutes? The answer is no. A twelve-case training set completed in less than 30 minutes. METARULE's run times depend as much on the internal structure of the data as on training set size.

Problem set two (figure 5) has four cases and completed in 12 minutes. METARULE again found the preset maximum of 50 rules, 17 of which correctly diagnose 100% of the cases (appendix B). The best scoring rule is:

A case is POSITIVE if:
the case has at least one white feature weighing 1
AND
the case doesn't have at least one green feature.

The last phrase could be better rendered as "the case doesn't have any green features". The awkwardness arises because the prototype structured English generator negates a phrase simply by preceding it with a negative word like "no" or "doesn't". The generator doesn't enable a negative to modify the target noun phrase, e.g. transform "has at least one" to "doesn't have any". This should be addressed in future versions.

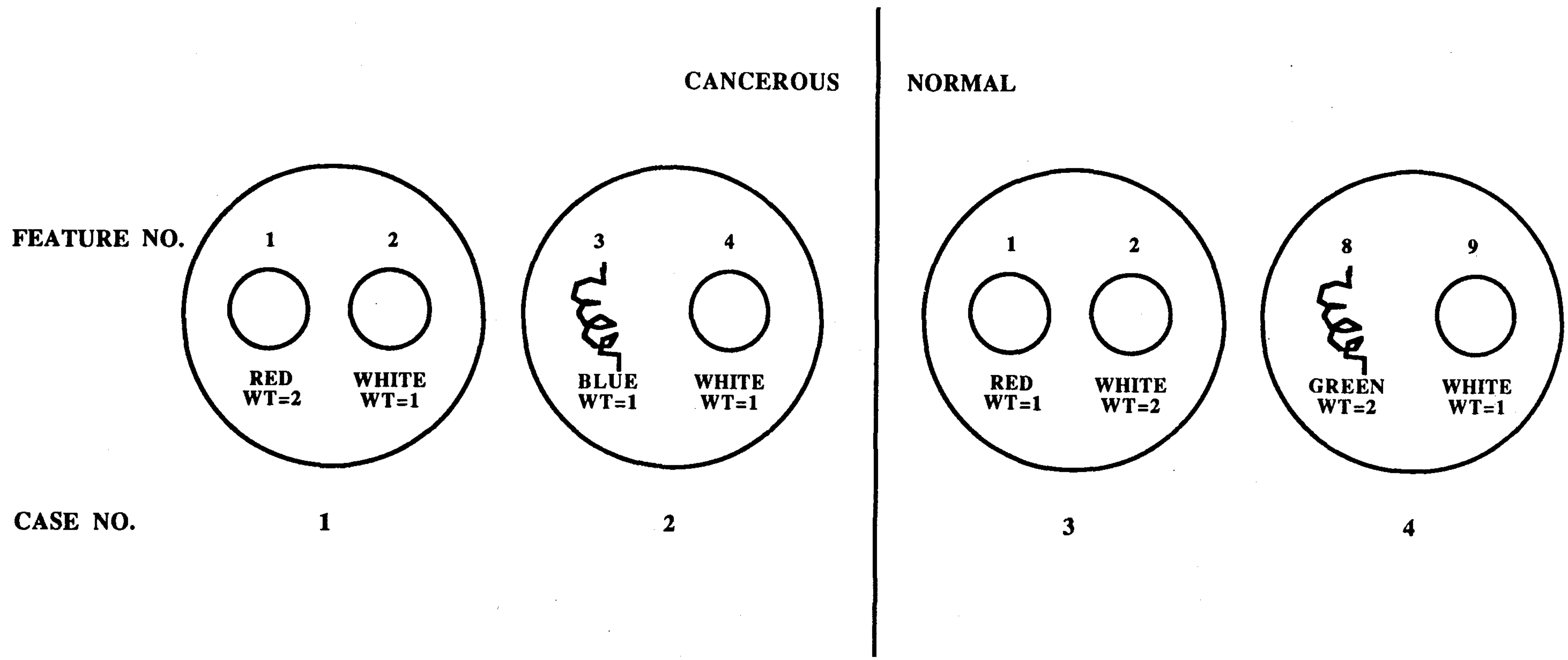

Figure 5. Test Problem Number Two:
What Makes a "Cell" Cancerous?

An alternative rule reads:

A case is POSITIVE if:
    the case has at least one red feature weighing 2
OR
        the case has at least one white feature weighing 1
    AND
        the case has exactly 1 blue feature.

The expert in Martian cells would select the simplest, clearest, most general and most valid rules for use, from among the 50 solutions.

In addition to presenting alternative solutions, METARULE also incorporates expert judgment by asking the user about regularities that it discovers in the data. In the domain of Martian cells, circles are the only normal feature and springs are the only abnormal feature. METARULE detects the consequences of this singularity and asks:

I have noticed that for the training set, saying that:
    There is at least one circle
is the same as saying:
    there is at least one normal feature.
Is this true in general (Y/N)?

If the user answers "no", METARULE performs induction on both circles and normal things. If the user answers "yes", METARULE asks which term the user prefers: circles or normal features. The feature which the user prefers is used for induction. The other feature is marked specially so that it is not used, but is reported as a variant to a rule.

For example, if the user prefers to talk in terms of circles, METARULE would append to each rule involving a circle a note that there are other ways to phrase the rule, and give the user the option to see them. The variants would substitute "normal features" for "circles".

It developed that this feature was undesirable for small training sets. The reader will perhaps have noted with surprise the large number of perfectly performing rules for sample problem 1, and discovered that the small size of the training set leaves a lot of possibilities open (figure 6). The same applies to the number of empirical equivalences: METARULE asked an inordinate number of questions. It remains to be determined whether or not this is a problem in really large training sets.

These sample problems are representative of the testing that occurred prior to receiving thallium data. To test METARULE, a set of physician's thallium worksheets were requested. The worksheets describe features of diagnostic significance, and thus mimic the output of the feature extractor.

What was in fact received was a page of draft rules for diagnosing images, together with a set of notional case data exemplifying those rules in application. The rules were in accord with the principles discussed in section 2. They were used to draft a set of 11 notional cases (table 2).

For the thallium domain, at least two runs are required for a training set. Like most induction systems, METARULE thinks of the cases as "positive" or "negative". But thallium permits three outcomes: "abnormal", "borderline" and "normal". In one run, "abnormal" cases are considered positive, and METARULE generates rules for diagnosing those cases. In the other run, "borderline" cases are considered "positive", and another set of rules is generated. Normal cases are those which are rejected by both rule sets.

1). FOR THE TRAINING SET, SAYING THAT

THE CASE HAS AT LEAST ONE NORMAL FEATURE

IS EQUIVALENT TO SAYING THAT

THE CASE HAS AT LEAST ONE CIRCLE.

IS THIS TRUE IN GENERAL?

2). FOR THE TRAINING SET, SAYING THAT

THE CASE HAS AT LEAST ONE FEATURE WEIGHING 2

IS THE SAME AS SAYING

THE CASE HAS AT LEAST ONE CIRCLE.

IS THIS TRUE IN GENERAL?

3). IN THE TRAINING SET, SAYING THAT

THE CASE HAS AT LEAST ONE WHITE FEATURE

IS THE SAME AS SAYING

THE CASE HAS AT LEAST ONE CIRCLE.

IS THIS TRUE IN GENERAL?

Figure 6. Three of METARULE's Questions for Sample Problem One

CASE NO.	CLASSIFICATION	FEATURES
1	NORMAL	NONE
2	NORMAL	REPERFUSION DEFECT POSTERIOR WALL THICKNESS .25 2 SUBSEGMENTS
3	NORMAL	MATCHED DEFECT NEAR VALVE PLANE
4	BORDERLINE	REPERFUSION DEFECT INFERIOR WALL THICKNESS .45 2 SUBSEGMENTS
5	NORMAL	REPERFUSION DEFECT NEAR VALVE PLANE THICKNESS .8 1 SUBSEGMENT
6	BORDERLINE	REPERFUSION DEFECT SEPTAL WALL THICKNESS .35 1 SUBSEGMENT
7	ABNORMAL	REPERFUSION DEFECT ANTERIOR WALL THICKNESS .75 1 SUBSEGMENT
8	ABNORMAL	REPERFUSION DEFECT POSTERIOR WALL THICKNESS .6 6 SUBSEGMENTS
9	ABNORMAL	REPERFUSION DEFECT INFERIOR WALL THICKNESS .6 2 SUBSEGMENTS
10	NORMAL	REPERFUSION DEFECT NEAR VALVE PLANE THICKNESS .75 1 SUBSEGMENT
11	NORMAL	REPERFUSION DEFECT NEAR VALVE PLANE THICKNESS .4 1 SUBSEGMENT

Table 2. Test Problem Number Three:
What Determines a "Thallium Image's" Classification?

Alternatively, one could make a third run on "normal" cases, and assign a diagnosis of "unknown" to imagery which meets none of the three rule sets.

Three runs were made. The best-scoring rules for NORMAL and BORDERLINE cases appear in figure 7.

From these and other sample problems, the following conclusions can be drawn about METARULE:

1). METARULE is capable of discovering rules from training sets having hierarchically organized nominal and numerical case data, like the data found in medical telemetry problems. The representational scheme and induction algorithm are appropriate to such problems.

2). METARULE's speed is acceptable. A several-fold increase in speed is expected when the prototype is ported to C. METARULE has not been timed on a really large training set. However, when generating a rule set, quality of result is more important than speed.

3). The speed and quality of inductive reasoning depend on the numerical parameters governing the induction process. The elaboration limit, branching factor and weighting factor must be adjusted to the characteristics of a particular training set.

4). The structured English generator is adequate for a prototype but requires refinement before productization.

5). The METARULE prototype demonstrates the feasibility of machine learning in the medical telemetry domain.

A case is **NORMAL** if
The case has no features with high location significance.

This solution correctly classifies 83.3% of the NORMAL cases and rejects 100% of the other cases. The performance rating of this rule is .917

A case is **BORDERLINE** if
The case has at least one reperfusion defect
AND
The case has at least one feature with thickness less than .6
AND
The case has at least one feature with high location significance.

This solution correctly classifies 100% of the NORMAL cases and rejects 100% of the other cases. The performance rating of this rule is 1.0.

A case is **ABNORMAL** if
There is at least one feature with high location significance
AND
There is no reperfusion defect with high location significance with 1 subsegment involved.

This solution correctly classifies 100% of the NORMAL cases and rejects 100% of the other cases. The performance rating of this rule is 1.0.

Figure 7. Best Scoring Rules for Test Problem Number Three.

BIBLIOGRAPHY

[ARV73] Arvenson, O. "A Hardware and Software System for the Automatic Evaluation of Electro- and Vectorcardiography" in Zywetz, C. and Schneider, R. (eds). Computer Applications in ECG and VCG Analysis (Amsterdam: North- Holland 1973).

[BRA85] Brachman, R. and Levesque, J (eds). Introduction to their book, Readings in Knowledge Representation (Los Altos CA: Morgan Kaufmann) 1985.

[BUC84] Buchanan, B. and Shortliffe, E. Rule-Based Expert Systems (Reading MA: Addison Wesley) 1984.

[COH86] Cohen, A. Biomedical Signal Processing (Boca Raton FL: CRC Press) 1986.

[FRO87] Froelicher, V. Exercise and the Heart. (Year Book Medical Publishers) 1987.

[GER87] Gerson, M. Cardiac Nuclear Medicine (Englewood Cliffs NJ: Prentice Hall) 1987.

[HOL80] Holman, B. et al. Practical Nuclear Cardiology. (Hagerstown MD: Harper and Row) 1979.

[ING75] Ingbar, G (ed.) Signal Analysis and Pattern Recognition in Biomedical Engineering (New York: Wiley) 1975.

[ISR80] Isreal, J. et al. "The Event-Related Brain Potential as an Index of Display-Monitoring Workload" Human Factors 22:2 1980.

[KAY88] Kay, Tom. Personal communication.

[LAI84] Laister, M. and Riggs, R. "A Comparison of Template Methods and Feature Extraction Methods of ECG Analysis" in Computers in Cardiology (IEEE) 1984.

[MAC85] McKeon, D. et al. "Rule-Based Interpretation of Aerial Imagery" IEEE Trans. Pattern Analysis and Machine Intelligence 7:5 1985.

[MIC83] Michelski, R. "A Theory and Methodology of Inductive Learning" in Michelski, R. et al. (eds.) Machine Learning: An Artificial Intelligence Approach. (Los Altos CA: Morgan Kaufmann) 1983.

[MIL84] Miller, P. A Critiquing Approach Expert Computer Advice (Boston: Pitman) 1984.

[PAT82] Patil, S. et al. "Modeling Knowledge of the Patient in Acid-base and Electrolyte Disorders" in Slovitz, P. (ed.) Artificial Intelligence in Medicine (AAAS) 1982.

[POE84] Poeppl, S. "Methods in One and Two Dimensional Signal Processing" in Fu, K. and Pavlidis, T (eds.) Biomedical Pattern Recognition (Berlin: Dahlem Konferenz) 1979.

[QUI68] Quillian, R. "Semantic Memory" in Minsky, M (ed.). Semantic Information Processing (Cambridge MA: MIT Press) 1968.

[RUM87] Rummelhart, D. and McClelland, J. Parallel Distributed Processing, volume 1. (MIT Press) 1986.

[SCH87] Schwartz, R. et al. "The Value of Exercise Thallium 201 Myocardial Scintigraphy for Detecting Coronary Artery Disease in Asymptomatic Men) USAFSAM 1987.

[SHO84] Shortliffe, E. et al. "An Expert System for Oncology Protocol Management" in [BUC84]

[SKL78] Sklansky, J. et al. "Boundary Detection in Medical Radiographs" in Preston, K. and Onone, M. (eds.) Digital Processing of Biomedical Images (New York: Plenum) 1976.

[SPR79] Sprengelmeyer, J. Practical Nuclear Cardiology. (Hagerstown MD: Harper & Row) 1979.

[THO86] Thompson, B. and Thompson, W. "Finding Rules in Data" Byte 11:12 1986.

[UTG86] Utgoff, P. Machine Learning of Inductive Bias (Boston: Elsevier) 1986.

[WAT81] Watson, J. et al. "Spatial and Temporal Quantification of Plane THallium Myocardial Images" J. Nuc. Med. 22:573 1981.

[WIN87] Winslow, B. Address at the First International Conference on Neural Nets (San Diego CA 1987).

[YU84] Yu, V. et al. "An Evaluation of MYCIN's Advice" in [BUC84].

Appendix A

METARULE Structured English Rules for Test Problem #1

Appendix A

METARULE Structured English Rules
for
Test Problem #1

1) A case is POSITIVE if:
   there are circles such that their combined characteristics include the following: white, weighing 1
   AND
   there are exactly two features in the case.

2) A case is POSITIVE if:
   there are circles such that their combined characteristics include the following: white, weighing 1
   AND
   the case has exactly one feature weighing 2.

3) A case is POSITIVE if:
   there are circles such that their combined characteristics include the following: white, weighing 1
   AND
   the case doesn't have exactly 1 white feature.

4) A case is POSITIVE if:

   there are circles such that their combined characteristics include the following: white, weighing 1
   AND
   the case doesn't have exactly 1 feature weighing 1.

5) A case is POSITIVE if:
   there are normal features such that their combined characteristics include the following: white, weighing 1
   AND
   there are exactly 2 features in the case.

6) A case is POSITIVE if:
   there are normal features such that their combined characteristics include the following: white, weighing 1
   AND
   the case has exactly 1 feature weighing 2.

7) A case is POSITIVE if:
   there are exactly two features in the case
   AND
   there are normal features such that the combined characteristics include the following: white, weighing 1.

Appendix A

METARULE Structured English Rules
for
Test Problem #1
(Continued)

8) A case is POSITIVE if:
   there are normal features such that their combined characteristics include the following: white, weighing 1
   AND
   the case doesn't have exactly 1 white feature.

9) A case is POSITIVE if:
   there are normal features such that their combined characteristics include the following: white, weighing 1
   AND
   the case doesn't have exactly 1 feature weighing 1.

10) A case is POSITIVE if:
   there are exactly two features in the case
   AND
   there are circles such that their combined characteristics include the following: white, weighing 1.

11) A case is POSITIVE if:
   there are normal features such that their combined characteristics include the following: white, weighing 1
   AND
   there are exactly 2 features in the case
   AND
   the case doesn't have exactly 1 white feature.

12) A case is POSITIVE if:
   there are normal features such that their combined characteristics include the following: white, weighing 1
   AND
   there are exactly 2 features in the case
   AND
   there is exactly 1 feature weighing 2.

13) A case is POSITIVE if:
   there are circles such that their combined characteristics include the following: white, weighing 1
   AND
   there are exactly 2 features in the case
   AND

the case does not have exactly 1 feature weighing 1.

14) A case is POSITIVE if:
there are circles such that their combined characteristics include the following: white, weighing 1
AND
there are exactly 2 features in the case
AND
the case doesn't have exactly 1 white feature.

15) A case is POSITIVE if:
there are circles such that their combined characteristics include the following: white, weighing 1
AND
there are exactly 2 features in the case
AND
the case has exactly 1 feature weighing 2.

16) A case is POSITIVE if:
there are circles such that their combined characteristics include the following: white, weighing 1
AND
there are normal features such that their combined characteristics include the following: white, weighing 1
AND
there are exactly 2 features in the case.

17) A case is POSITIVE if:
there are circles such that their combined characteristics include the following: white, weighing 1
AND
there are normal features such that their combined characteristics include the following: white, weighing 1
AND
the case doesn't have exactly 1 feature weighing 1.

18) A case is POSITIVE if:
there are circles such that their combined characteristics include the following: white, weighing 1
AND
there are normal features such that their combined characteristics include the following: white, weighing 1

AND
the case doesn't have exactly 1 white feature.

19) A case is POSITIVE if:
there are circles such that their combined characteristics include the following: white, weighing 1
AND
there are normal features such that their combined characteristics include the following: white, weighing 1
AND
the case has exactly 1 feature weighing 2.

20) A case is POSITIVE if:
there are normal features such that their combined characteristics include the following: white, weighing 1
AND
there are exactly 2 features in the case
AND
the case doesn't have exactly 1 feature weighing 1.

21) A case is POSITIVE if:
there are circles such that their combined characteristics include the following: white, weighing 1
AND
there are normal features such that their combined characteristics include the following: white, weighing 1
AND
there are exactly 2 features in the case
AND
the case doesn't have exactly 1 feature weighing 1.

22) A case is POSITIVE if:
there are circles such that their combined characteristics include the following: white, weighing 1
AND
there are normal features such that their combined characteristics include the following: white, weighing 1
AND
there are exactly 2 features in the case
AND
the case doesn't have exactly 1 white feature.

Appendix A

METARULE Structured English Rules
for
Test Problem #1
(Continued)

Appendix A

METARULE Structured English Rules
for
Test Problem #1
(Continued)

23) A case is POSITIVE if:
there are circles such that their combined characteristics include the following: white, weighing 1
AND
there are normal features such that their combined characteristics include the following: white, weighing 1
AND
there are exactly 2 features in the case
AND
the case has exactly 1 feature weighing 2.

24) A case is POSITIVE if:
there are circles such that their combined characteristics include the following: red, weighing 2
OR
there are circles such that their combined characteristics include: white, weighing 1.
AND
the case has exactly 1 circle.

25) A case is POSITIVE if:
there are circles such that their combined characteristics include the following: red, weighing 2
OR
there are circles such that their combined characteristics include: white, weighing 1.
AND
the case has exactly 1 normal feature.

26) A case is POSITIVE if:
there are circles such that their combined characteristics include the following: red, weighing 2
OR
there are circles such that their combined characteristics include: white, weighing 1.
AND
the case has no red feature.

Appendix A

METARULE Structured English Rules
for
Test Problem #1
(Continued)

27) A case is POSITIVE if:
there are circles such that their combined characteristics include the following: red, weighing 2

OR

there are circles such that their combined characteristics include: white, weighing 1.

AND

the case has no red feature weighing 2.

28) A case is POSITIVE if:
there are circles such that their combined characteristics include the following: red, weighing 2

OR

there are circles such that their combined characteristics include: white, weighing 1.

AND

there is some feature type with features whose characteristics include the following: red, weighing 2.

29) A case is POSITIVE if:
there are circles such that their combined characteristics include the following: red, weighing 2

OR

there are circles such that their combined characteristics include: white, weighing 1.

AND

there is some feature class with features whose characteristics include the following: red, weighing 2.

30) A case is POSITIVE if:
there are circles such that their combined characteristics include the following: red, weighing 2

OR

there are circles such that their combined characteristics include: white, weighing 1.

AND

the case doesn't have at least one red feature.

Appendix A

METARULE Structured English Rules
for
Test Problem #1
(Continued)

31) A case is POSITIVE if:
there are circles such that their combined characteristics include the following: red, weighing 2

OR

there are circles such that their combined characteristics include: white, weighing 1.

AND

the case doesn't have exactly 1 green feature.

32) A case is POSITIVE if:
there are circles such that their combined characteristics include the following: red, weighing 2

OR

there are circles such that their combined characteristics include: white, weighing 1.

AND

the case doesn't have exactly 1 spring.

33) A case is POSITIVE if:
there are circles such that their combined characteristics include the following: red, weighing 2

OR

there are circles such that their combined characteristics include: white, weighing 1.

AND

the case doesn't have exactly 1 abnormal feature.

34) A case is POSITIVE if:
there are circles such that their combined characteristics include the following: red, weighing 2

OR

there are normal features such that their combined characteristics include: white, weighing 1.

AND

the case has exactly 1 circle.

Appendix A

METARULE Structured English Rules
for
Test Problem #1
(Continued)

35) A case is POSITIVE if:
there are circles such that their combined characteristics include the following: red, weighing 2
OR
there are normal features circles such that their combined characteristics include: white, weighing 1.
AND
the case has exactly 1 normal feature.

36) A case is POSITIVE if:
there are circles such that their combined characteristics include the following: red, weighing 2
OR
there are normal features such that their combined characteristics include: white, weighing 1.
AND
the case has no red feature.

37) A case is POSITIVE if:
there are circles such that their combined characteristics include the following: red, weighing 2
OR
there are normal features such that their combined characteristics include: white, weighing 1.
AND
the case has no red feature weighing 2.

38) A case is POSITIVE if:
there are normal features such that their combined characteristics include the following: red, weighing 2
OR
there are circles such that their combined characteristics include: white, weighing 1.
AND
the case has no red feature weighing 2.

Appendix A

METARULE Structured English Rules for Test Problem #1 (Continued)

39) A case is POSITIVE if:
there are normal features such that their combined characteristics include the following: red, weighing 2
OR
there are circles such that their combined characteristics include: white, weighing 1.
AND
the case has exactly 1 normal feature.

40) A case is POSITIVE if:
there are normal features such that their combined characteristics include the following: red, weighing 2
OR
there are circles such that their combined characteristics include: white, weighing 1.
AND
the case has no red feature.

41) A case is POSITIVE if:
there are normal features such that their combined characteristics include the following: red, weighing 2
OR
there are circles such that their combined characteristics include: white, weighing 1.
AND
the case has no red feature weighing 2.

42) A case is POSITIVE if:
there are normal features such that their combined characteristics include the following: red, weighing 2
OR
there are circles such that their combined characteristics include: white, weighing 1.
AND
there is some feature type with features whose characteristics include the following: red, weighing 2.

Appendix A

METARULE Structured English Rules
for
Test Problem #1
(Continued)

43) A case is POSITIVE if:
there are normal features such that their combined characteristics include the following: red, weighing 2
OR
there are circles such that their combined characteristics include: white, weighing 1.
AND
there is some feature class with features whose characteristics include the following: red, weighing 2.

44) A case is POSITIVE if:
there are normal features such that their combined characteristics include the following: red, weighing 2
OR
there are circles such that their combined characteristics include: white, weighing 1.
AND
the case doesn't have at least one red feature.

45) A case is POSITIVE if:
there are normal features such that their combined characteristics include the following: red, weighing 2
OR
there are circles such that their combined characteristics include: white, weighing 1.
AND
the case doesn't have exactly one green feature.

46) A case is POSITIVE if:
there are normal features such that their combined characteristics include the following: red, weighing 2
OR
there are circles such that their combined characteristics include: white, weighing 1.
AND
the case doesn't have exactly one spring.

METARULE Structured English Rules for Test Problem #1 (Continued)

47) A case is POSITIVE if:
there are normal features such that their combined characteristics include the following: red, weighing 2
OR
there are circles such that their combined characteristics include: white, weighing 1.
AND
the case doesn't have exactly one abnormal feature.

48) A case is POSITIVE if:
there are normal features such that their combined characteristics include the following: red, weighing 2
OR
there are normal features such that their combined characteristics include: white, weighing 1.
AND
the case has exactly 1 circle.

49) A case is POSITIVE if:
there are normal features such that their combined characteristics include the following: red, weighing 2
OR
there are normal features such that their combined characteristics include: white, weighing 1.
AND
the case has exactly 1 normal feature.

50) A case is POSITIVE if:
there are normal features such that their combined characteristics include the following: red, weighing 2
OR
there are normal features such that their combined characteristics include: white, weighing 1.
AND
the case has no red feature.

Appendix B

METARULE Structured English Rules
for
Test Problem #2

Appendix B

METARULE Structured English Rules
for
Test Problem #2

1) A case is POSITIVE if:
   the case has at least one white feature weighing 1
   AND
   the case doesn't have at least one green feature.

2) A case is POSITIVE if:
   there is some feature type with features whose characteristics include the following: white, weighing 1
   AND
   the case doesn't have at least one green feature.

3) A case is POSITIVE if:
   there are circles such that their combined characteristics include the following: white, weighing 1
   AND
   the case doesn't have exactly 1 green feature.

4) A case is POSITIVE if:

   the case has at least one red feature weighing 2
   OR
   the case has at least one blue feature.

5) A case is POSITIVE if:
   the case has at least one red feature weighing 2
   OR
   the case has exactly one blue feature.

6) A case is POSITIVE if:
   the case has at least one red feature weighing 2
   OR
   the case has at least one white feature weighing 1
   AND
   the case has exactly 1 blue feature.

7) A case is POSITIVE if:
   the case has at least one red feature weighing 2
   OR
   there are circles such that their combined characteristics include the following: white, weighing 1
   AND
   the case has exactly 1 blue feature.

Appendix B

METARULE Structured English Rules
for
Test Problem #2
(Continued)

8) A case is POSITIVE if:
   there are circles such that their combined characteristics include the following: red, weighing 2
   OR
      the case has at least one blue feature.

9) A case is POSITIVE if:
   there are circles such that their combined characteristics include the following: red, weighing 2
   OR
      the case has exactly one blue feature.

10) A case is POSITIVE if:
   there are circles such that their combined characteristics include the following: red, weighing 2
   OR
      the case has at least one white feature weighing 1
   AND
      the case has exactly one blue feature.

11) A case is POSITIVE if:
   there are circles such that their combined characteristics include the following: red, weighing 2
   OR
      there are circles such that their combined characteristics include the following: white, weighing 1
   AND
      the case has exactly one blue feature.

12) A case is POSITIVE if:
   there is some feature type with features whose characteristics include the following: red, weighing 2
   OR
      the case has at least one blue feature.

13) A case is POSITIVE if:
   there is some feature type with features whose characteristics include the following: red, weighing 2
   OR
      the case has exactly 1 blue feature.

Appendix B

METARULE Structured English Rules for Test Problem #2 (Continued)

14) A case is POSITIVE if:
    there is some feature type with features whose characteristics include the following: red, weighing 2
  OR
        the case has at least one white feature weighing 1
    AND
        the case has exactly 1 blue feature.

15) A case is POSITIVE if:
    there is some feature type with features whose characteristics include the following: red, weighing 2
  OR
        there are circles such that their combined characteristics include the following: white, weighing 1
    AND
        the case has exactly 1 blue feature.

16) A case is POSITIVE if:
    there is some feature type with features whose characteristics include the following: red, weighing 2
  OR
        the case has at least one blue feature.

17) A case is POSITIVE if:
    there is some feature type with features whose characteristics include the following: red, weighing 2
  OR
        the case has exactly 1 blue feature.